

# Architecture flexible pour la stéréovision embarquée

Mehdi Darouich <sup>(1)</sup>, Stéphane Guyetant <sup>(1)</sup> et Dominique Lavenier <sup>(2)</sup>

<sup>(1)</sup> CEA, LIST, PC94, 91191 Gif sur Yvette

<sup>(2)</sup> IRISA, Campus de Beaulieu, 35042 Rennes

## I. INTRODUCTION

La stéréovision permet l'extraction de la profondeur d'une scène à partir des images de deux capteurs vidéo. Dans le domaine automobile, elle est principalement utilisée pour la détection et la localisation d'obstacles. Les avantages apportés par la stéréovision peuvent bénéficier à plusieurs applications d'aide à la conduite (ou Advanced Driver Assistance System, ADAS) comme la détection de piétons, la mesure de distance de sécurité, l'adaptation des phares à la situation et à la topologie ou la détection de panneau de signalisation. Outre les aspects de sûreté, ces applications doivent répondre à de fortes contraintes du temps-réel et de l'embarqué. Pour la plupart des traitements mis en jeu, un compromis doit être fait entre la qualité des résultats et le besoin en ressources. Notre objectif est de proposer une solution architecturale assurant la meilleure efficacité énergétique tout en garantissant la qualité du résultat. Nous nous intéressons plus particulièrement à la création de la carte de disparité qui permet l'extraction de la profondeur et qui est une étape très calculatoire. L'objectif ici est d'obtenir une carte de disparité dense et d'extraire le plan de la route de manière robuste.

## II. PRINCIPE DE LA DISPARITÉ

La disparité est la différence de position entre un point de l'image de la première caméra et son correspondant sur l'image de la deuxième caméra. L'information de disparité  $d$ , couplée aux paramètres physiques de la paire stéréoscopique - base de la paire  $B$  et distance focale des caméras  $f$  - permet d'obtenir la profondeur  $P$  du point considéré selon la formule  $P = B.f/d$ .



Fig. 1. Carte de disparité et résultat de la détection de piéton

L'objectif de l'estimation de disparité est de trouver la correspondance entre les points des images de droite et de gauche. Une rectification préalable des deux images permet de limiter le champ de recherche : les deux points correspondants se trouvent sur la même ligne des deux images. Malgré

cette simplification, cette phase de recherche est hautement calculatoire et la qualité des résultats influe fortement sur les performances des traitements en aval comme la segmentation et la classification. Le plus souvent, un compromis doit être fait entre le temps de calcul et la qualité de l'image de disparité (densité et niveau d'erreur).

## III. ÉTUDE ALGORITHMIQUE

Les algorithmes de mise en correspondance peuvent être classés selon trois familles : méthodes locales, globales et semi-globales. Les méthodes locales basent leur analyse sur une fenêtre de comparaison autour des pixels à mettre en correspondance. Dans la zone de recherche, un coût d'appariement est obtenu par corrélation entre les deux fenêtres. Les méthodes les plus couramment employées sont la somme des écarts quadratiques (SSD), la somme des écarts absolus (SAD) ou l'intercorrélation normalisée centrée (ZNCC). Par la suite, la disparité pour laquelle on obtient le meilleur score d'appariement est retenue (approche Winner-Take-All). Ces méthodes simples présentent l'avantage de nécessiter le moins de ressources et de générer des images de disparité denses, mais elles présentent un fort taux d'erreurs, notamment dans les zones d'occlusion et dans les zones peu texturées - par exemple pour une route neuve.

D'autres approches, regroupées sous le nom de méthodes globales, consistent à optimiser une fonction d'énergie définie sur toute l'image. La fonction d'énergie impose les contraintes auxquelles la carte de disparité doit répondre : par exemple la continuité de la disparité sur les objets. Par la suite, l'ensemble des disparités qui minimisent cette fonction d'énergie est recherché. La méthode de coupure de graphes (Graph-Cut, GC) et la propagation de croyance (Belief Propagation, BP) sont les deux méthodes globales les plus étudiées. Ces méthodes donnent des images de disparité denses et comportant très peu d'erreurs. Elles sont cependant complexes à mettre en place et nécessitent des ressources de calcul et de mémorisation très importantes, incompatibles avec les contraintes matérielles de l'embarqué (cf. tableau I).

Les méthodes semi-globales se basent sur le même principe que les méthodes globales mais sur des sous-parties de l'image (des lignes ou des blocs). Le découpage du problème d'optimisation en sous-problèmes permet de diminuer le besoin en ressources de calcul et en ressources mémoires par rapport aux méthodes globales, mais entraîne cependant l'apparition d'artefacts sur l'image de disparité. Une méthode

de propagation de croyance appliquée à des blocs de pixels (Block-BP) [1] permet la mise en place d'un traitement en parallèle. L'optimisation de la disparité suivant les lignes épipolaires peut s'effectuer par programmation dynamique (DP). Dans cette méthode, l'objectif est de trouver le chemin qui minimise la fonction d'énergie à travers l'espace des disparités possibles sur une ligne (Disparity Space Image ou DSI) dont la représentation est donnée figure 2. La programmation dynamique extrait l'ensemble des disparités donnant la meilleure corrélation tout en respectant la contrainte de continuité, ce qui permet l'obtention de cartes de disparité denses, homogènes, mais présentant des défauts sur les contours qui peuvent déformer certains objets.

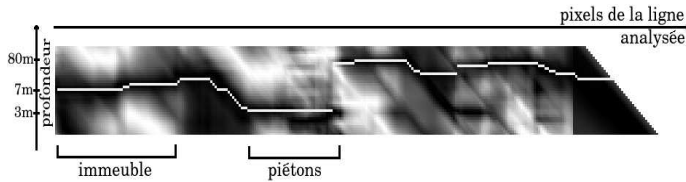


Fig. 2. Image d'espace de disparité (DSI) de la ligne marquée d'une flèche figure 1. Plus la corrélation est élevée, plus le pixel du DSI est sombre. Le chemin blanc dans les zones sombres représente la disparité extraite par programmation dynamique

Nous avons implémenté et testé les principaux algorithmes sur des paires d'images stéréoscopiques de situations routières. Plusieurs métriques ont été introduites pour évaluer ces algorithmes. Le besoin en ressources est mesuré par le nombre d'opérations (normalisées en 8 bits) nécessaires par seconde. L'occupation mémoire représente la taille minimale du buffer nécessaire au traitement. La bande passante mémoire est le débit d'information entre une mémoire externe et le buffer précédemment défini. Les résultats de notre étude sont regroupés dans le tableau I. Les images en niveau de gris 8 bits ont une résolution de 640x480, la fréquence des images à traiter est de 40 fps, le niveau de disparité maximum est de 60, et la fenêtre de traitement a pour taille 5x5.

TABLE I  
ÉVALUATION DES ALGORITHMES DE CRÉATION D'IMAGE DE DISPARITÉ

		Besoin en ressources (GOPS)	Occupation mémoire	Bande passante mémoire
Méthodes locales	SAD	56	7ko	135Mo/s
	ZNCC	225	7ko	135Mo/s
Méthodes globales	BP	71541	442Mo	N/A
Méthodes semi-globales	DP	74	182ko	135Mo/s
	Block-BP	2120	737ko	36,9Mo/s

#### IV. ARCHITECTURES POUR ADAS

Il existe plusieurs architectures matérielles visant les applications d'aide à la conduite. L'architecture EyeQ2 de Mobileye [2], qui sera commercialisé début 2009, associe deux processeurs RISC à virgule flottante MIPS34K, trois processeurs

vectoriels, cinq blocs de traitement dédiés, dont un pour le calcul de la disparité et 1Mo de mémoire SRAM. Cette solution permet d'implémenter un grand nombre d'applications de vision. Cependant, l'utilisation de blocs de traitement dédiés ne permet pas d'évolution dans les traitements qu'ils adressent, et la flexibilité réside seulement dans la programmation des processeurs.

IMAP-CE [3] est une architecture SIMD composée de 128 processeurs VLIW à 4 voies 8 bits possédant chacun 2ko de buffer mémoire. Elle peut fournir 51,2 GOPS de performance crête en exploitant le parallélisme de données, pour une consommation de 2W. La version commerciale, IMAPCAR, permet d'obtenir de meilleures performances, environ 100 GOPS, en intégrant des opérateurs 16 bits. L'implémentation d'un algorithme basé sur la programmation dynamique a été effectué sur la version précédente, IMAP-VISION, constituée de 256 processeurs RISC 8 bits avec 1ko de mémoire chacun. Le système génère l'image de disparité dense sur 64 niveaux d'une région d'intérêt de 128x128 pixels en 100ms.

#### V. CONCLUSION

Ce poster présente notre démarche de conception d'une architecture flexible pour la stéréovision appliquée aux systèmes d'aide à la conduite. L'étude algorithmique des différentes solutions de création de carte de disparité nous a permis d'extraire les besoins en ressources, mémoire et bande passante pour ce type d'applications. Ces mesures et le niveau de performances qu'offrent les architectures existantes montrent que, dans le cas d'images de haute résolution ou des méthodes semi-globales, les contraintes temps-réels fortes sont difficilement atteignables. Une solution possible serait l'augmentation du nombre d'éléments de calculs et de la taille mémoire, impliquant cependant une surface silicium plus importante. Les solutions à base d'accélérateurs matériels possèdent une excellente performance au détriment de la flexibilité algorithmique.

Nous considérons l'intégration d'éléments de calcul reconfigurables comme une solution pour augmenter cette flexibilité et c'est cette approche que nous visons dans la suite de l'étude. L'extraction des opérations nécessaires nous permettra de définir une ou plusieurs briques élémentaires reconfigurables ainsi qu'une topologie mémoire adaptée. Enfin, la validation de l'architecture nécessite la définition d'une chaîne algorithmique adaptée aux contraintes routières et bornée en complexité calculatoire.

#### REFERENCES

- [1] Y.-C. Tseng, N. Chang, and T.-S. Chang, "Low memory cost block-based belief propagation for stereo correspondence," *Multimedia and Expo, 2007 IEEE International Conference on*, pp. 1415–1418, 2–5 July 2007.
- [2] MOBILEYE, "http://www.mobileye.com."
- [3] S. Okazaki, "An integrated memory array processor for embedded image recognition systems," *IEEE Trans. Comput.*, vol. 56, no. 5, pp. 622–634, 2007, member-Shorin Kyo and Member-Tamio Arai.